IIIORRISON FOERSTER

425 MARKET STREET
SAN FRANCISCO
CALIFORNIA 94105-2482

TELEPHONE: 415.268.7000
FACSIMILE: 415.268.7522

WWW.MOFO.COM

MORRISON & FOERSTER LLP

AUSTIN, BEIJING, BERLIN, BOSTON, BRUSSELS, DENVER, HONG KONG, LONDON, LOS ANGELES, MIAMI, NEW YORK, PALO ALTO, SAN DIEGO, SAN FRANCISCO, SHANGHAI, SINGAPORE, TOKYO, WASHINGTON, D.C.

February 27, 2024

Writer's Direct Contact
+1 (415) 268-6066
JGratz@mofo.com

VIA ECF

Hon. Sidney H. Stein
Daniel Patrick Moynihan
United States Courthouse
500 Pearl St.
New York, NY 10007-1312

Re: *Authors Guild et al. v. OpenAI, Inc., et al.*, and *Alter et al. v. OpenAI Inc., et al.*,
   Nos. 1:23-cv-08292-SHS & 1:23-cv-10211-SHS

Dear Judge Stein:

Pursuant to Rule 2(G) of Your Honor's Individual Practices, Defendant OpenAI submits this response to Plaintiffs' February 23 request for an informal discovery conference concerning Plaintiffs' Requests for Admissions ("RFAs"). Plaintiff has served nearly a thousand RFAs seeking to learn new information about what works do and do not appear in the training data for OpenAI's models—often asking OpenAI to admit both that a particular work is part of the training data and that it isn't. Further, some of the information that Plaintiffs seek is irrelevant because Plaintiffs have served RFAs that, on their face, ask about *all* OpenAI models, not just the ones at issue in this case. Because Plaintiffs' requests are improper and unduly burdensome under Federal Rule of Civil Procedure 36, Plaintiffs' request for a discovery conference should be denied.

First, Plaintiffs' RFAs are for the purpose of learning new information, not confirming facts that they believe are not disputed: Plaintiffs do not know whether certain books appear in the training datasets for OpenAI's models, and they wish to learn this information through the service of hundreds of RFAs. But the federal rules do not allow that. OpenAI is not required to answer hundreds of RFAs that have been served "to elicit facts and information rather than to confirm information." *See Spectrum Dynamics Med. Ltd. v. General Elec. Co*, 18-cv-11386 (VSB) (KHP), 2021 WL 735241, at *2 (S.D.N.Y. Feb. 25, 2021). That is because RFAs "are particularly ill-suited to the task of uncovering new information and thus generally are best used at the conclusion of discovery, because their sole purpose is to streamline the presentation of evidence at trial." *BAT LLC v. TD Bank, N.A.*, No. 15-cv-5839 (RRM) (CLP), 2018 WL 3626428, at *6 (E.D.N.Y. July 30, 2018) (quoting *T. Rowe Price Small-Cap Fund, Inc. v. Oppenheimer & Co.*, 174 F.R.D. 38, 43 (S.D.N.Y. 1997)).

IIIORRISON FOERSTER

Hon. Sidney H. Stein
February 27, 2024
Page Two

Accordingly, courts in this district have often recognized that Rule 36 is not a discovery tool. To discover new information, parties must use other methods, like depositions, document requests, or interrogatories. *See, e.g.*, *Republic of Turkey v. Christie's, Inc.*, 326 F.R.D. 394, 399 (S.D.N.Y. 2018) (explaining that "[w]hile the basic purpose of discovery is to elicit facts and information and to obtain production of documents, Rule 36 was not designed for this purpose") (quoting 7 *Moore's Federal Practice* § 36.02[1]); *Spectrum Dynamics*, 2021 WL 735241, at *2 (explaining that RFAs "are not designed to discover information like other discovery rules such as Rule 34" and excusing a party from responding where RFAs were "tantamount to contention interrogatories"). If an RFA raises issues that are "more appropriately addressed through deposition and document discovery," courts have held that a response is not required. *Carver v. Bank of New York Mellon*, 15-CV-10180 (JPO) (JLC), 2018 WL 4579831, at *2 (S.D.N.Y. Sept. 25, 2018).

Here, Plaintiffs have served 939 requests concerning an issue that should be "addressed through deposition and document discovery." *See Carver*, 2018 WL 4579831, at *2. Through these requests, Plaintiffs seek to elicit new information about the contents of the datasets used to train OpenAI's models (as demonstrated by the fact that Plaintiffs have served RFAs asking Plaintiffs both to admit that certain works appear in the training data and to admit that those same works do not appear). This is a paradigmatic attempt to *discover* information about OpenAI, rather than an effort to *confirm* information that Plaintiffs already know. Accordingly, Plaintiffs must use a more appropriate, and less burdensome, discovery device. *See BAT*, 2018 WL 3626428, at *6 (excusing responses to RFAs where "it appears that rather than seeking admissions regarding facts already known to plaintiff, [the] requests for admissions actually seek to discover information that is central to the dispute in this action and not already known" to the requesting party). Indeed, OpenAI has already agreed to make training data available for inspection as soon as is practicable. Inspecting that data, rather than serving nearly a thousand RFAs, is the appropriate way for Plaintiffs to learn the information they seek.

Second, even if they were legally proper, the RFAs are unduly burdensome both because of the number of RFAs and because Plaintiffs have not provided OpenAI with copies of the works at issue. OpenAI's training datasets are voluminous. It is not realistic to expect OpenAI to devise and run the searches necessary to respond to all 939 requests at the outset of this case, especially without the benefit of copies of the works at issue. *See BAT*, 2018 WL 3626428, at *6 (finding 226 RFAs to be "overly burdensome and not proportional to the needs of the case at this stage of the litigation"). GPT-4, for example, was developed using, among other things, information that is publicly available on the internet. Was some or all of the text of Jonathan Alter's book *The Center Holds: Obama and His Enemies* somewhere within that data, as a result of having been posted publicly on the internet somewhere? The

IIIORRISON FOERSTER

Hon. Sidney H. Stein
February 27, 2024
Page Three

only way to find out is to search for passages from the text of that book in the dataset, and that is what Plaintiffs should do as part of their inspection.

Third, even if OpenAI would have to answer RFAs as to the models used for ChatGPT, Plaintiffs' requests are not so limited. Instead, they ask about *all* of OpenAI's large language models—including those that have nothing to do with ChatGPT. Plaintiffs' claims, however, are about ChatGPT, and ChatGPT alone. (*See, e.g.*, First Consolidated Complaint (ECF No. 67) ¶¶ 103-108, 143-153, 176-180, 186-190, 196-199, 205-209, 215-219, 225-229, 235-239, 244-247, 253-257, 263-265, 270-274, 279-285, 291-295, 301-305, 311-315, 321-325, 331-335, 341-345, 349, 352, 355, 359, 363, 367, 370, 374, 378, 382, 386, 390.) OpenAI should not, therefore, have to provide information about models that were never used for ChatGPT.

In conclusion, OpenAI seeks to cooperate with Plaintiffs to provide access to the appropriately discoverable information that they seek. To that end, OpenAI is working diligently to make training datasets available for inspection as soon as is practicable, and OpenAI told Plaintiffs as much when the parties met and conferred by phone on February 6. Through inspection, Plaintiffs will be able to determine for themselves which works do and do not appear in that data. Accordingly, a Rule 34 inspection is the superior—and legally required—way for Plaintiffs to elicit information about the contents of OpenAI's training datasets. OpenAI should not be required to respond to Plaintiffs' hundreds of RFAs.

Sincerely,

/s/ Joseph C. Gratz

Joseph C. Gratz
Morrison & Foerster LLP

/s/ Joseph R. Wetzel

Joseph R. Wetzel
Latham & Watkins LLP

cc: All Counsel of Record